AWARD NUMBER: W81XWH-11-1-0401

TITLE: Mammary Cancer and Activation of Transposable Elements

PRINCIPAL INVESTIGATOR: Anne Peaston

CONTRACTING ORGANIZATION: The University of Adelaide
Adelaide, 5000 Australia

REPORT DATE: March 2015

TYPE OF REPORT: Final

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

DISTRIBUTION STATEMENT: Approved for Public Release;
Distribution Unlimited

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
March 2015	Final	01 Sep 2011 - 31 Dec 2014

4. TITLE AND SUBTITLE

Mammary Cancer and Activation of Transposable elements

5a. CONTRACT NUMBER
W81XWH-11-1-0401

5b. GRANT NUMBER
BC101296

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)

Anne Peaston

E-Mail:anne.peaston@adelaide.edu.au

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

The University of Adelaide
Adelaide, 5000 Australia

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)

U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT

Approved for Public Release; Distribution Unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

The purpose of this project is to investigate molecular events of the preclinical stages of mammary cancer, specifically, the intersection between the development of genome demethylation, retrotransposon transcriptional activity, and retrotransposon-driven transcription of cellular genes in an engineered parity-dependent mouse model of mammary cancer. These inbred mice develop mammary cancer after 3 litters. Mammary epithelial cells (MEC) were isolated from cancer-prone and control mice after 1 or 3 litters. DNA and RNA were isolated from the MEC. Data from Illumina HiSeq RNA libraries was forwarded to our collaborator Dr John Edwards for transcriptome analysis. DNA from MEC was sent to Dr Edwards for processing and methylome analysis. As expected, using Methyl-MAPS analysis, Dr Edwards identified whole genome hypomethylation in tumor-prone MEC compared with controls after 3 litters. We did not identify increased retrotransposon expression. An unexpected and intriguing observation was upregulated expression of immune system suppressors in tumor-prone MEC after 3 litters. Wet lab independent confirmation of the findings is pending, and forms part of the research plan for a doctoral student in Adelaide. Our data suggest that while hypomethylation is an early event in this model, upregulated retrotransposon expression is a later event, and unlikely to play a direct role in cancer ontogeny.

15. SUBJECT TERMS

Breast cancer, epigenetic, DNA methylation, retrotransposon, preclinical cancer development, deep sequencing

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			USAMRMC
U	U	U	UU	14	19b. TELEPHONE NUMBER *(include area code)*

Table of Contents

1. Introduction

This project is designed to address the subject of mammary cancer development. The purpose of the project is to investigate molecular events occurring in the preclinical stages of mammary cancer; the results may lead to insights into cancer prevention in the future. Specifically, the project investigates the intersection between genome demethylation, retrotransposon transcriptional activity, and retrotransposon-driven transcription of cellular genes. Retrotransposon promoters are well recognized to function as alternative promoters for different cellular genes, generating chimeric transcripts that may or may not function in the same way as transcripts from the regular gene promoter. Transcriptional activation of retrotransposons is strongly linked with their CpG DNA methylation, and global genomic demethylation is one of the commonest molecular changes in malignancies. The project tests the hypothesis that, in preclinical stages of tumour development, progressive genomic demethylation leads to increased transcriptional activity of retrotransposons and this, in turn, leads to transcription of otherwise silent genes, potentially setting up molecular conditions that favour cancer development. We developed a genetically engineered mouse model in which a specific mammary cell population is fluorescently marked upon initial transcriptional activation of the SV40 large T antigen (SV40Tag) oncogene. SV40Tag protein disables p53 and Rb proteins, mimicking a common molecular pathology in human cancers. In Waptag1 mice, SV40Tag is transcriptionally activated during pregnancy and lactation, and the mice are predisposed to develop mammary cancer after a minimum of 3 pregnancies and lactations. Using this model, we collected populations of marked cells for integrated analysis of gene expression, promoter usage, and DNA methylation after defined amounts of exposure to SV40Tag during different stages of preclinical cancer development.

2. Keywords

Cancer
Cancer development
Mammary cancer
Mouse model
Retrotransposon
Alternative promoters
DNA methylation
SV40 large T antigen
Methyl-MAPS

3. Overall Project Summary

Introduction.

We are interested in the very earliest events that set the scene for mammary cancer development, with a view to eventual application to cancer prevention. Under the influence of the Whey Acidic Protein promoter (Wap) Waptag1 mice express SV40Tag in mammary epithelial cells (MEC) only during pregnancy and lactation. After 3 pregnancies and lactations, but not after 1 pregnancy and lactation, females develop mammary cancers at an average age of about 13 months. This suggests that, in a long-lived MEC population, successive exposures to SV40Tag create molecular effects that eventually reach a threshold permitting cancer to develop. We hypothesized that genome-wide hypomethylation would be a preclinical event in this cell population, and that a consequence would be indiscriminate activation of retrotransposons. We predicted that this in turn would result in deranged cellular gene expression as a result of retrotransposon promoter-driven expression of otherwise silent cellular genes, potentially creating a molecular milieu favoring cancer development.

In preparation for this project, while at The Jackson Laboratory Dr Peaston developed a line of mice congenic on the C57BL6/J background, expressing Cre recombinase driven by the Wap promoter (WapCre mice) and a two-color fluorescent Cre constitutive reporter allele in the ROSA locus ($ROSA^{mT/mG}$). The mTmG;WapCre double transgenic mice constitutively express red fluorescence in cells prior to Cre recombinase exposure. In the first pregnancy and lactation, MEC that express WapCre switch permanently to expression of enhanced green fluorescent protein (EGFP). No other cell type in the mammary fat pad was observed to express EGFP. Wholemount and FACS analyses of mammary fat pads after involution from lactation showed that a population of EGFP-expressing cells, or parity-induced MEC (PI-MEC), persisted in duct structures. When crossed with

Waptag1 mice (mTmG;WapCre;Waptag1), mammary tumours that developed in triparous F1 triple hemizygotes consisted of EGFP-expressing cells, indicating that the cells were descendents of the long-lived PI-MEC marked in the first pregnancy by EGFP expression. Thus, cancer-prone cells could be readily marked for isolation and molecular investigations in preclinical stages of cancer development.

Project design

The project was originally set up as a collaborative effort between Dr Edwards (Washington University in St Louis; methylome analysis) and Dr Peaston (The Jackson Laboratory, Bar Harbor, Maine; animal experiments, RNA and DNA isolation and processing) with key personnel for bioinformatics analyses from The Jackson Laboratory. Dr Peaston transferred the project to the University of Adelaide following her academic appointment at that institution in 2011. Participation of The Jackson Laboratory was discontinued on departure of key personnel, and Dr Edwards took over all aspects of the bioinformatics analysis as well as methyl DNA library construction and analysis. CAGE analysis was eventually abandoned because of lower than anticipated RNA yields. These changes are outlined in Figure 1.

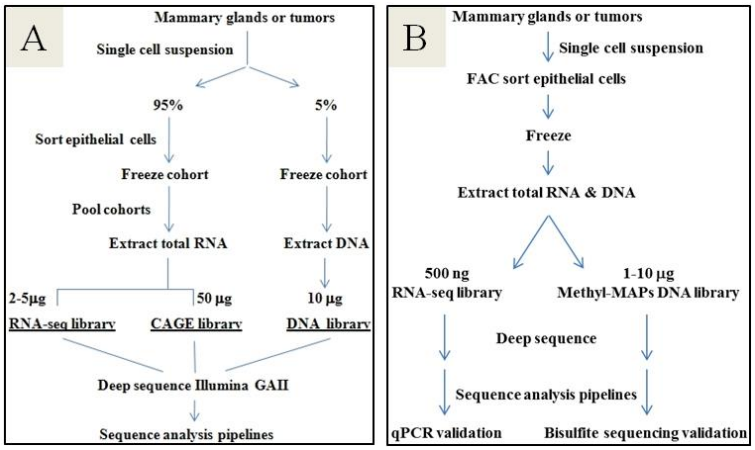

Figure 1. Project flow charts. A. Original plan to be conducted between University of Adelaide, The Jackson Laboratory and Washington University B. Revised plan conducted between University of Adelaide and Washington University.

After importation and establishment of the mouse breeding colonies in Adelaide, Dr Peaston instigated the

A

	Cohort					
Week	1	2	3	4	5	6
1	12 trios*					
2						
3		12 trios				
4						
5	pups		12 trios			
6						
7		pups		12 trios		
8						
9	wean/mate		pups		12 trios	
10						
11		wean/mate		pups		12 trios
12						
13	pups		wean/mate		pups	
14						
15		pups		wean		pups
16	wean/mate					
17			pups		wean	
18		wean/mate				
19				kill		wean
20	final wean		wean/mate			
21					kill	
22		final wean				
23						kill
24	kill		final wean			
25						
26		kill				
27						
28			kill			

B

mating	Mating cohort							
week	1	2	3	4	5	6	7	8
1	mate*	Initial 12-14 cages trio matings						
2								
3		mate*						
4	pups							
5			mate*	Switch to 24-28 cages pair matings				
6		pups						
7	wean			mate*				
8			pups					
9		wean			mate*			
10				pups				
11	mate		wean			mate*		
12					pups			
13		mate		wean			mate*	
14	pups					pups		
15			mate		wean			mate*
16		pups					pups	
17	wean			mate		wean		
18			pups					pups
19		wean			kill		wean	
20				pups				
21	mate		wean			kill		wean
22								
23		mate		wean			kill	
24	pups							
25			mate					kill
26		pups						
27	wean			mate				
28			pups					
29		wean						
30				pups				
31	kill		wean					
32								
33		kill		wean				
34								
35			kill					
36								
37				kill				

Figure 2. Mouse breeding schematic for experimental females, one biologic replicate each for 3 and 1 pregnancy/lactations. Three biologic replicates were undertaken. A. Original plan. B. plan modified to accommodate 48-60 mated females per experimental condition (1 or 3 pregnancies/lactations.

breeding strategy to develop triple transgenic cancer-prone and control experimental female mice (mTmG;WapCre;Waptag1 SVTag/+ and SV+/+ respectively). To investigate molecular features of preclinical cancer development, EGFP-expressing PI-MEC were isolated from SVTag/+ and SV+/+ mice that had undergone one pregnancy and lactation (uniparous) or three pregnancies and lactations (triparous). At least 4 weeks after weaning each litter, allowing for complete involution of the mammary glands, mice were either mated for the next litter or sacrificed for PI-MEC isolation and processing. Figure 2 shows the initial plan for mouse husbandry, with staggered cohorts of experimental animals. The biologic replicates for one and three pregnancy/lactations were alternated to minimise cumulative mouse cage numbers. The major modification was to initiate pair matings instead of trio matings after the first two cohorts of mice in an attempt to improve the poorer than expected reproductive performance, at the expense of extending the calculated period of experiment by 3 months. Some later cohorts had exceptionally poor reproductive success with as few as 17% pregnant in the allotted two week period, necessitating further extension of the breeding period of the project; the reasons for the poor breeding performance were not definitively identified. Over the course of the project, the proportion of mated mice that successfully completed their planned pregnancies and were sacrificed for PI-MEC extraction was relatively poor, 53% for 1 pregnancy and lactation, either genotype, or 46% (SVTag/+) and 54% (SV+/+) for 3 pregnancies and lactations (Table 1).

Table 1. Summary of Experimental Mouse Usage

			One pregnancy		Three pregnancies		TOTAL
		genotype	SV/+	+/+	SV/+	+/+	
Biological Replicate	1	mated	72	79	49	63	263
		culled*	44	41	26	34	145
		sacrificed**	28	38	23	29	118
	2	mated	50	77	57	56	240
		culled*	19	41	28	20	108
		sacrificed**	31	36	29	36	132
	3	mated	54	67	49	79	249
		culled*	20	22	30	38	110
		sacrificed**	34	45	19	41	139
TOTAL		mated	176	223	155	198	752
		culled*	83	104	84	92	363
		sacrificed**	93	119	71	106	389

Matings for 48-60 mice per genotype per replicate were planned for each condition.

* mice euthanased or died without mammary collection for reasons such as poor reproductive performance (n=157), tumour development (soft tissue sarcoma, lymphoma n=59); death due to unknown reason (n=46), and smaller numbers of animals with various conditions (malocclusion, head tilt, dystocia, respiratory complaints, identity not clear, hydrocephalus, hydronephrosis, infection, injury, paraplegia, prolapsed rectum or uterus, seizures. ** mice killed to harvest PI-MEC.

Because of facility limitations, mice were sacrificed for PI-MEC isolation in groups of up to 4 control or cancer-prone uniparous or triparous females. Both 4th mammary fat pads were removed from sacrificed mice and a commercially available kit was used according to the manufacturer's instructions (EasySep™ Mouse Epithelial Cell Enrichment Kit, StemCell Technologies) to prepare single cell suspensions through enzymatic and physical methods, and to enrich for MEC by immunomagnetic depletion of CD45+, CD31+, TER119+, BP-1+ cells. The cells from each group of mice were pooled and pure EGFP-expressing PI-MEC populations were isolated by fluorescence-activated cell sorting (FACS) (Figure 3). EGFP-expressing PI-MEC pooled cell pellets were snap frozen or were resuspended in minimum quantity of lysis buffer (AllPrep DNA/RNA Micro Kit, Qiagen) before freezing for later extraction of RNA and DNA.

RNA and DNA were isolated from individual frozen cell pools, or from two to three matched pools combined, (Qiagen AllPrep DNA/RNA Micro Kit), and quantitated using a Nanodrop apparatus. RNA yields were persistently lower than expected, likely due to a combination of factors centering on, but not limited to,

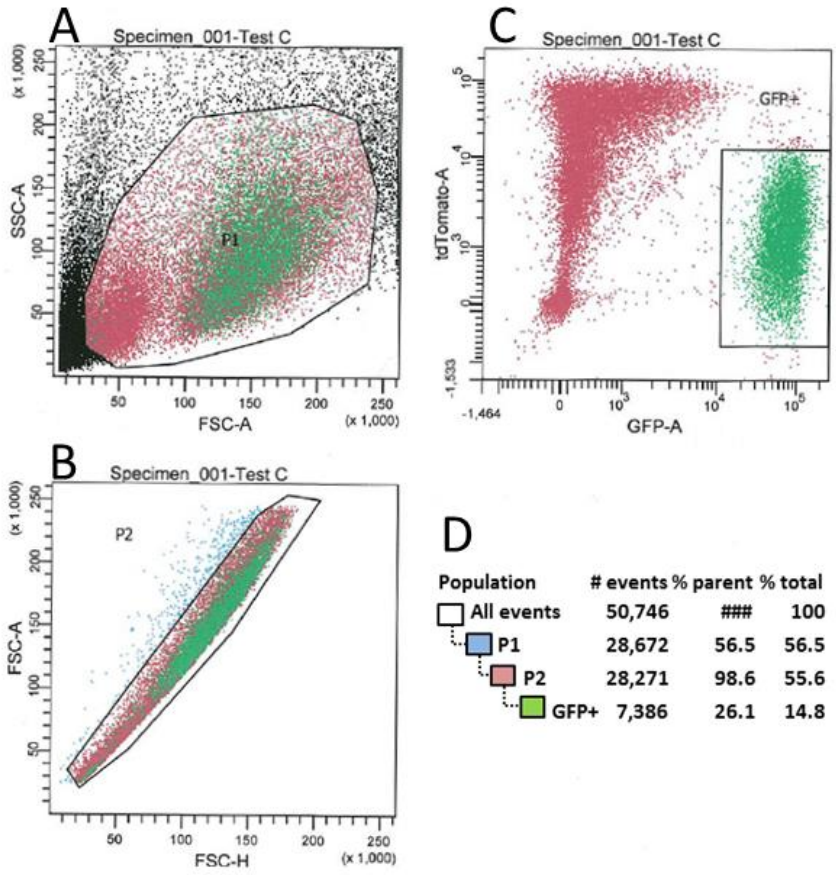

Figure 3. FACS gating strategy for resolving PI-MEC populations. Single cell suspensions (CD45$^-$, CD31$^-$, TER119$^-$, BP-1$^-$) prepared from uniparous or triparous SV/+ or SV+/+ mice were pooled for collection of PI-MEC. A: Dead and dying cells collect in a band along the side of the side scatter (SSC-A) versus forward scatter (FSC-A) plot, and these were gated out to define P1. B: cell doublets were discarded to define P2 C: The target cell population of PI-MEC fluoresces bright green, forming a discrete subpopulation of P2. D: Gating tree showing gating strategy for FACS, and analysis of cell percentages within the gates. The total sorted EGFP cell count in this experiment was 1.2 x 10^6, but the intact cell yield after FACS was less than half. Data is from a representative pool of cells from 4 triparous SVTag/+ mice.

Table 2. Biologic replicate samples of RNA for library construction, extracted from tumour-prone (SVTag/+) or control (SV+/+) PI-MEC in uniparous or triparous mice.

Parity	SV40Tag genotype	RNA ng/µl	RNA volume µl	Total RNA ng
1	SVTag/+	6	29	174
	SVTag/+	8	29	232
	SV+/+	17	29	493
	SV+/+	17	29	493
	SV+/+	4	29	116
3	SVTag/+	6	29	174
	SVTag/+	3	29	87
	SV+/+	16	29	464
	SV+/+	17	29	493

unavoidable cell loss during FACS in combination with the modest numbers of EGFP-positive PI-MEC in the cell suspensions (Fig3 D, Table 2).

Altogether, 46 SV+/+ (21 uniparous and 25 triparous) and 31 SVTag/+ (19 uniparous, 12 triparous) RNA and DNA extractions were prepared for molecular investigations. For expression analysis, 23 RNA preparations representing cells from uniparous or triparous tumour-prone or control mice from each biologic replicate were sent to the Australian Genome Research Facility (http://www.agrf.org.au) but only 9 samples passed Agilent Bioanalyser quality assurance as suitable for RNA library preparation (Table 2). These

samples allowed library preparation from two but not the original three planned biologic replicates of the experiment. Illumina 100 bp paired-end RNA libraries were prepared and sequenced in triplicate from the 9 samples. Sequencing data were generated with Illumina CASAVA pipeline version 1.8.2, and the data yield is illustrated in Table 3. Expression library data was forwarded to Dr Edwards for transcriptome analysis.

Table 3: 100bp paired end data yield. Source:AGRF report 26 June 2014

Lane	Parity	SV40Tag genotype	Replicate	Paired end	Data yield (bp)
3	1	SVTag/+	1	21,154,786	4.23 Gb
		SVTag/+	2	20,870,347	4.17 Gb
		SV+/+	1	21,239,715	4.25 Gb
		SV+/+	2	20,932,118	4.19 Gb
		SV+/+	3	21,366,538	4.27 Gb
	3	SVTag/+	1	22,012,694	4.40 Gb
		SVTag/+	3	21,145,553	4.23 Gb
		SV+/+	2	17,675,342	3.54 Gb
		SV+/+	3	22,097,649	4.42 Gb
4	1	SVTag/+	1	21,320,575	4.26 Gb
		SVTag/+	2	21,031,255	4.21 Gb
		SV+/+	1	21,398,576	4.28 Gb
		SV+/+	2	21,116,707	4.22 Gb
		SV+/+	3	21,534,095	4.31 Gb
	3	SVTag/+	1	22,182,913	4.44 Gb
		SVTag/+	3	21,287,743	4.26 Gb
		SV+/+	2	17,797,107	3.56 Gb
		SV+/+	3	22,269,182	4.45 Gb
5	1	SVTag/+	1	21,059,778	4.21 Gb
		SVTag/+	2	20,796,624	4.16 Gb
		SV+/+	1	21,158,829	4.23 Gb
		SV+/+	2	20,852,642	4.17 Gb
		SV+/+	3	21,264,104	4.25 Gb
	3	SVTag/+	1	21,914,698	4.38 Gb
		SVTag/+	3	21,080,912	4.22 Gb
		SV+/+	2	17,598,050	3.52 Gb
		SV+/+	3	22,017,374	4.40 Gb
Total				**566,175,906**	**113.24 Gb**

DNA isolated from pooled PI-MEC from 2-4 mice was forwarded to Dr Edwards for generation of Methyl-MAPS whole genome methylome data, and integrated analysis with the RNA library data. There was insufficient RNA to conduct Cap-Associated Gene Expression (CAGE) promoter analyses, and the remainder of the RNA has been held for real-time quantitative PCR validation studies of the expression data.

Bioinformatics analysis

Over the course of the project, Dr Edwards has developed and refined new computational tools to examine the relationship between genome-wide DNA methylation and gene expression. Previous approaches from many different groups illuminated the genomic organization of CpG methylation marks, but were unsatisfactory for evaluating how changes at specific loci affected function. The WIMSi (Washington University Methylation Signatures) pipeline enables Dr Edwards to combine genome-wide methylation and expression data to develop a high-resolution view of the relationship between DNA methylation and specific transcript production. An

initial description of Dr Edwards' methods and results has been published and the full manuscript is in the appendix to his final report for this project (*1*).

Dr Edwards' pipeline permits unbiased genome-wide DNA profiling to reveal previously unknown information, identifying genes whose methylation and expression change in a pairwise fashion. The tool was designed to start from a list of expression data, corresponding transcription start sites (TSSs) and high-resolution genome-wide methylation data such as from Methyl-MAPS. Thus it fits perfectly into the framework of this proposal where we have obtained RNA-seq expression and Methyl-MAPS methylation data for each sample. Dr Edwards further adapted the tool to address the regulation of retrotransposons promoters and has performed some initial analyses of LTR-driven lncRNAs as a successful proof-of-concept.

Expression analysis of MECs from uni- and triparous mice.

Expression analyses have been conducted in Dr Edwards' laboratory, and much of the data reported here also appear in his report. RNA-Seq analyses using long-paired end reads were performed for two biologic replicates. Each replicate consists of four samples, or each combination of uni- or triparous, and SV+/+ or SVTag/+ mice. For each sample, 30-46 million paired reads were aligned to the mouse genome (mm10) using Tophat (*2*). Mapped reads were aligned to Ensembl genes using htseq and differential expression was computed using edgeR (*3*). Genes that were at least 2-fold up or down regulated with FDR (False Discovery Rate) < 0.05 were deemed significant. 318 genes were uniquely down and 34 uniquely up in tumor-prone triparous mice. Biologic process analysis of the 318 down regulated genes using web-based software DAVID version 6.7 (*4, 5*) indicated that genes involved in the cell cycle and downregulation of the immune response were over-represented in this group (Figure 4). Genes regulating cell cycle control could represent the early inactivation of checkpoints that create a permissive environment for tumor proliferation. Genes involved in immune response could represent exaggerated attenuation of the immune response that accompanies involution. Functional enrichment was also analysed by PANTHER Overrepresentation Test (release 20150430) (http://geneontology.org/, PANTHER version 10.0 released 2015-05-15) and the results were in broad agreement with the DAVID 6.7 analysis. We are currently scouring the literature for clues to understanding how these genes may be involved in setting the regulatory environment to facilitate tumor development, and to prioritise quantitative PCR experiments on independent samples to validate the expression data. Also intriguing in the DAVID 6.7 analysis was enrichment for genes involved in inititation of apoptosis (data not shown). Whether this reflects a link with prior inhibition of Tp53 and associated apoptotic pathways by

Figure 4. Screen shot of the top 2 and part of the 3rd clusters of functional enrichment categories (biologic process) of transcripts downregulated in PI-MEC from triparous SVTag/+ compared with triparous SV+/+ mice. Analysis: DAVID 6.7 https://david.ncifcrf.gov/home.jsp, high stringency. Numbers of genes contributing to the cluster are shown in the Count column.

Annotation Cluster 1	Enrichment Score: 9.9	Count	P_Value	Benjamini
GOTERM_BP_FAT	regulation of T cell activation	17	4.5E-12	6.7E-9
GOTERM_BP_FAT	regulation of cell activation	18	1.4E-10	5.4E-8
GOTERM_BP_FAT	regulation of lymphocyte activation	17	3.8E-10	1.1E-7
GOTERM_BP_FAT	regulation of leukocyte activation	17	1.0E-9	1.7E-7
Annotation Cluster 2	**Enrichment Score: 8.55**	**Count**	**P_Value**	**Benjamini**
GOTERM_BP_FAT	mitotic cell cycle	22	8.1E-11	4.1E-8
GOTERM_BP_FAT	M phase of mitotic cell cycle	19	5.8E-10	1.4E-7
GOTERM_BP_FAT	cell cycle phase	24	5.9E-10	1.1E-7
GOTERM_BP_FAT	M phase	22	1.3E-9	1.9E-7
GOTERM_BP_FAT	nuclear division	18	3.1E-9	3.9E-7
GOTERM_BP_FAT	mitosis	18	3.1E-9	3.9E-7
GOTERM_BP_FAT	organelle fission	18	5.4E-9	6.2E-7
GOTERM_BP_FAT	cell cycle process	24	1.9E-8	1.7E-6
GOTERM_BP_FAT	cell cycle	30	2.6E-8	2.0E-6
GOTERM_BP_FAT	cell division	20	3.6E-8	2.4E-6
Annotation Cluster 3	**Enrichment Score: 6.5**	**Count**	**P_Value**	**Benjamini**
GOTERM_BP_FAT	positive regulation of T cell activation	12	3.0E-9	4.1E-7
GOTERM_BP_FAT	regulation of lymphocyte differentiation	11	1.1E-8	1.1E-6
GOTERM_BP_FAT	positive regulation of T cell differentiation	9	1.5E-8	1.5E-6
GOTERM_BP_FAT	positive regulation of lymphocyte differentiation	9	2.6E-8	2.0E-6
GOTERM_BP_FAT	regulation of T cell differentiation	10	2.9E-8	2.1E-6
GOTERM_BP_FAT	positive regulation of cell activation	13	3.0E-8	2.0E-6
GOTERM_BP_FAT	positive regulation of lymphocyte activation	12	1.3E-7	8.4E-6
GOTERM_BP_FAT	positive regulation of leukocyte activation	12	2.3E-7	1.4E-5
GOTERM_BP_FAT	positive regulation of immune system process	16	4.1E-7	2.3E-5

SV40Tag is not clear. Genes involved in apoptosis were not flagged as significantly overrepresented in the PANTHER Overrepresentation Test, probably because the different software programs use different ontology databases and different methods of analysis. .

Analyses of chimeric transcripts (containing both retroelement and coding sequence) are underway. To our surprise, when we analyzed retrotransposon-derived sequences we found their expression dropped in tumor-prone triparous mice by an average of ~18%. This occurred across nearly every transposable element family. We are currently working to validate these findings.

Methyl-MAPS analysis of MECs from uni- and triparous mice.
Dr Edwards has performed Methyl-MAPS analysis for one experimental replicate and has presented the results in detail in his report. Difficulties were encountered during the final stages of library construction for replicate 2, and attempts to salvage these samples and perform Methyl-MAPS analysis are under way. For replicate 3 it was determined that there was insufficient material to build sequencing libraries and thus the DNA has been held for validation of target regions identified from the other replicates.

In brief, the data indicate no significant methylation level difference between control and tumour-prone MEC from uniparous mice. However, in PI-MEC from triparous mice the tumor-prone cells show an approximately 3.5% drop in methylation compared with cells from control triparous mice (Figure). These data support a hypothesis that mechanisms driving the large-scale genomic hypomethylation observed in breast tumors may be active in preclinical stages of tumour development.

Integrated Methylation and Expression Analyses
These analyses have been reported in detail in Dr Edwards' report. Essentially, transcription activity was inversely correlated with methylation in the region 500 bp upstream to 1 kb downstream of transcription start sites and was positively correlated with methylation of gene bodies, in line with previous reports (*1, 6, 7*). However, there appeared to be little correlation between promoter methylation and expression levels of either coding genes or retrotransposons. It also appears that the promoter methylation changes that are common in primary tumors are not found in this early stage. We conclude that the mechanisms that cause demethylation may be active and may create a permissive environment for later transcriptional changes as tumors arise. However, these methylation changes do not appear to affect transcription of retrotransposons nor coding genes at this stage. In the future, we will characterize methylation and expression from tumors arising in tumor-prone triparous mice to further understand when the large-scale remodeling of DNA methylation occurs and how these methylation changes facilitate transcriptome remodeling.

Statement of Work

Year 1 & 2: Items from Statement of Work Relevant to Peaston Lab. Comprehensive final summary as at 31 March 2015. As detailed in the annual reports from 2012, and 2013, numerous technical difficulties and delays affected timely completion, and a no-cost extension was granted for 2014 ennabling completion of the most time-consuming tasks. As previously noted, we were unable to collect sufficient RNA to conduct CAGE experiments, and we are in the process of finalizing analyses and performing validation of hypotheses arising from the expression and methylation data. We hope to have a manuscript prepared and submitted detailing our findings of the analysis of uniparous and tumour-prone triparous mice within the next 6 – 9 months.

Year and months	GOAL	Result
YEAR 1		
preliminary activity	1. Hire Research Assistant 2. Import mice 3. – 5. Set up foundation & starter breeding colonies 6. Travel to USA/organise e-meeting schedules	• All completed
Months 1-3	1. Establish production colonies	• Completed
	2. – 3. collect pilot tissue samples from virgin mice and if possible from tumours, extract RNA/DNA, ship to collaborators.	• Incomplete – breeding problems dictated that there were no spare virgin or aging experimental mice for much of the project, and no tumours occurred in any mice over this period
	4. schedule meetings	• Completed in Goal 6 above and hereafter
Months 4-6	1. Initiate setting up cohorts for mouse breeding experiments	• Completed
	2-8 Collect & process experimental mammary samples and tumour samples if available, extract RNA/DNA, ship materials to collaborators, preliminary bioinformatics analyses	• Tumour collection planned for Waptag1 females was not done due to mismatch in tumour availability and processing arrangements. Funded otherwise, triparous tumor-prone mice were set aside in late 2014 for tumor development and molecular analyses, and virgin mice are being produced for analysis.
Months 7-9	1. Continue setting up cohorts for mouse breeding experiments	• Complete
	2. – 3. Experimental replicate #1 uniparous mice harvest cells, extract RNA/DNA, Ship materials to collaborators	• Complete
	3. set aside RNA for CAGE	• incomplete, insufficient to proceed with analysis
	4. Start RT-PCR evaluation of interesting transcripts from initial library analyses	• Not completed, underway in Dr Peaston's lab.
Months 10 -12	1. – 4, 7. Continue mouse breeding experiments, cell collection, nucleic acid isolation, shipping material to collaborators, RNA library prep & sequencing	• Complete.
	5. RNA for CAGE/preliminary analysis	• incomplete
	6. RT-PCR evaluation of interesting transcripts from initial library analyses	• Not completed, underway in Dr Peaston's lab.
YEAR 2		
Months 1-3	1. – 3. Complete mouse breeding experiments, collect cells, ship material to collaborators	• Complete
	4. Preliminary integration CAGE and expression data	• Incomplete. Preliminary material not available (see year 1 months 1-3 above) • CAGE analyses eventually abandoned due to insufficient RNA
	6. RT-PCR evaluation of interesting transcripts from initial library analyses	• Not completed, underway in Dr Peaston's lab.
	7. Bisulfite sequencing DNA loci of interest	• Not completed, under way in lab
Months 4 - 6	1. – 2. Complete mouse breeding experiments, prepare RNA/DNA, ship material to collaborators	• Complete (except for CAGE material, see above)
	4. – 5. RNA library prep & sequencing	• Complete
	6. RT-PCR evaluation of interesting transcripts from initial library analyses	• Not completed, underway in Dr Peaston's lab.

	7. Bisulfite sequencing DNA loci of interest	• Not completed, under way in lab
Months 7 - 9	1. – 3,6. Complete mouse breeding experiments, prepare RNA/DNA, ship material to collaborators, RNA library prep & sequencing, & analysis	• Complete
	5. Extra CAGE library	• Abandoned – insufficient RNA.
	7. RT-PCR evaluation of interesting transcripts from initial library analyses	• Not completed, underway in Dr Peaston's lab.
	8. Bisulfite sequencing DNA loci of interest	• Not completed, under way in lab
	10. – 11. Data integration and approaches to publishing.	• Under way between Peaston and Edwards labs.
Months 10-12	1. – 5. Final RT-PCR/bisulfite expts, data analysis, preparation of report for publication; ongoing contact between Edwards and Peaston labs.	• Continuing

4. Key Research Accomplishments

- Dr Edwards developed new computational tools to combine genome-wide expression and methylation data to output a list of genes where methylation likely contributes to their silencing or activation.

- We performed genome-wide methylation and expression profiling of SV40+/- and control uni- and tri-parous mice.

- We found a small amount of genomic demethylation specific to tumor-prone triparous mice that was not accompanied with corresponding changes in the transcriptome. These changes indicate that the mechanisms driving demethylation in breast tumors may have activated, even though they may not yet contribute to transcriptome remodeling at this early stage.

- We found downregulation of genes involved in immune system activation in PI-MEC from triparous tumour-prone mice. If confirmed, this result could indicate that local suppression of the immune-system is one of the earliest events in mammary tumourigenesis.

- Preliminary expression analyses suggest that demethylation may be uncoupled from retrotransposon activation in this setting, suggesting that retrotransposon transcription derangement is not a significant contributor to the earliest steps of tumourigenesis.

5. Conclusion

In retrospect, the work planned was overly ambitious for a 2 year project, and was fraught with persistent problems related to animal reproduction and sample preparation. The large number of experimental female mice committed to the project was specifically calculated to enable collection of the large quantity of RNA that CAGE analysis requires. Preliminary experiments at The Jackson Laboratory, where mouse housing, laboratory space, flow cytometry and sequencing facilities are under the one roof, supported the feasibility of the project. These conditions were not replicated in Adelaide and, in the long run, this likely contributed to the sample preparation problems, and ultimately the cancellation of the CAGE analysis project component. Nonetheless, we have successfully generated useful data from our expression and methyl-DNA studies, and are continuing to work on them.

The major scientific conclusion from this work is that at this early stage of tumor development, we find only a small amount of genome-wide hypomethylation in tumor-prone triparous mice and do not observe widespread DNA methylation changes that are commonly found in tumors. Correlated methylation changes at promoters and expression changes are found at hundreds of genes in primary human tumors, but none were observed in the tumor-prone mice in this study. Likely the mechanisms that lead to the genome-wide hypomethylation observed

in primary tumors have activated, but they are not impacting transcription of retrotransposons nor coding genes at this early stage.

Expression analyses indicate that the tumor-prone triparous mice experience altered transcription of several hundred transcripts. Many of these transcripts are found in suppression of the immune system and could represent failure to terminate immune suppression that follows involution. Many other transcripts are found in cell cycle control and could represent the early inactivation of checkpoints to control proliferation. Alternatively, since Tp53 and Rb are inactivated in PI-MEC by SV40Tag during the proliferative phases of pregnancy and lactation, a population of cells with unrepaired DNA defects may remain in the mammary gland after involution and be suppressed from proliferative activity by the regained checkpoint activities of Rb and Tp53. There does not appear to be wide-spread methylation or expression changes in retrotransposon sequences, indicating that these likely play little to no role in this stage.

During the course of this proposal we have developed a new set of computation tools that potentially have broad significance beyond the work here. The computational tools we have developed are designed to work with annotated genes as we have outlined, but can also be expanded to any transcriptional unit with a known TSS and known expression value. Further we have demonstrated how they can be directly used to study which genes are correlated, and thus potentially regulated, by DNA methylation in cancer cells. We have also shown how we can use these tools to identify genes that are potentially up-regulated in direct response to demethylating agents such as 5-azacitidine and Decitabine. While DNA demethylating agents have primarily been employed to treat Myelodysplastic syndrome and AML, recently these drugs have been explored as potential therapeutics in solid-tumors such as lung and breast cancer. We believe this work has the potential to shed light on which patients have genes that can be potentially re-activated by these drugs, and thus which patients can potentially benefit from this line of therapy.

6. Publications, Abstracts, and Presentations (not including Dr Edwards publications)

Abstracts

- Smith T., Edwards J., Peaston AE. Mammary Cancer and the Activation of Transposable Elements. Poster presentation at Australian Society for Medical Research, SA Division, Annual Scientific Meeting April 2014.
- Smith T., Edwards J., Peaston AE. Mammary Cancer and the Activation of Transposable Elements. Poster presentation at Australian Society for Medical Research, SA Division, Annual Scientific Meeting May 2015

Invited seminar

- Seminar “Retrotransposons and mammary cancer” The Basil Hetzel Institute, Queen Elizabeth Hospital, Adelaide. November 23, 2011.

7. Inventions, Patents and Licenses

Nothing to report

8. Reportable Outcomes

Nothing to reports

9. Other Achievements

- Australian Postgraduate Award received by Tim Smith for work continuing this project
- Manuscript in preparation (Tim Smith), Genome Methylation, Retrotransposition-driven Gene Expression And Mammary Cancer, a Review. In preparation for submission to *Cancer Science*

10. References

1. N. D. Vanderkraats, J. F. Hiken, K. F. Decker, J. R. Edwards, Discovering high-resolution patterns of differential DNA methylation that correlate with gene expression changes. *Nucleic acids research* **41**, 6816-6827 (2013).
2. C. Trapnell, L. Pachter, S. L. Salzberg, TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* **25**, 1105-1111 (2009).
3. M. D. Robinson, D. J. McCarthy, G. K. Smyth, edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139-140 (2010).
4. W. Huang da, B. T. Sherman, R. A. Lempicki, Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic acids research* **37**, 1-13 (2009).
5. W. Huang da, B. T. Sherman, R. A. Lempicki, Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature protocols* **4**, 44-57 (2009).
6. X. Yang *et al.*, Gene body methylation can alter gene expression and is a therapeutic target in cancer. *Cancer cell* **26**, 577-590 (2014).
7. R. A. Irizarry *et al.*, The human colon cancer methylome shows similar hypo- and hypermethylation at conserved tissue-specific CpG island shores. *Nat Genet* **41**, 178-186 (2009).

11. Appendices N/A